PubChem Laboratory Chemical Safety Summary

Sunghwan Kim*†, Leah R. McEwen‡, Ralph B. Stuart§, Paul A. Thiessen†, Asta Gindulyte†, Jian Zhang†, Evan E. Bolton†, Stephen H. Bryant†
† National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, 20894.
‡ Clark Physical Sciences Library, Cornell University, Ithaca, NY, 14853.
§ Office of Environmental Health & Safety, Keene State College, Keene, NH, 03435.

ABSTRACT

A series of high profile laboratory accidents over the last seven years has increased the academic community's awareness of the importance of chemical risk assessment in laboratories, both teaching and research. Because of the many different chemicals used in laboratory work, these risk assessments require ready access to safety information for a large number of chemicals. In order to provide researchers, educators and students with electronic access to chemical health and safety information, PubChem provides a "Laboratory Chemical Safety Summary" (LCSS) data view for thousands of chemicals commonly encountered in laboratories (https://pubchem.ncbi.nlm.nih.gov/lcss/).

The LCSS contains pertinent chemical health and safety data for a given PubChem Compound record. This data is organized based on the format described by the National Research Council in "*Prudent Practices in the Laboratory: Handling and Management of Chemical Hazards*". Information contained in the LCSS is a subset of the PubChem Compound Summary page content, and includes information on flammability, toxicity, exposure limits, exposure symptoms, first aid, handling, spill clean up and more. These data are provided by various data contributors, including authoritative agencies such as the U.S. National Institute for Occupational Safety and Health (NIOSH), the Occupational Safety and Health Administration (OSHA), and the International Labor Organization (ILO). Of special interest in this collection is specific "Stability and Reactivity" information, which lists known chemical reactions between the chemical of interest and other chemicals and chemical classes. This reactivity information often extends beyond that found on Safety Data Sheets for the chemical of interest.

LCSSs are available for PubChem Compound records which have GHS (Globally Harmonized System of Classification and Labeling of Chemicals) hazard classification information from PubChem's data sources. PubChem is actively recruiting this GHS information from additional contributors in order to expand its LCSS coverage. If a PubChem Compound record has an LCSS, the link to view it is provided at the top of the Compound Summary page of that compound under the heading "Safety Summary". In addition, one can get the complete list of chemicals with an LCSS by visiting the PubChem LCSS project webpage (https://pubchem.ncbi.nlm.nih.gov/lcss/) or by using the PubChem Classification Browser (https://pubchem.ncbi.nlm.nih.gov/classification/). If desired, LCSS data can be downloaded from the LCSS page for each compound, or in bulk from the PubChem LCSS project webpage.

INTRODUCTION

A series of high profile laboratory accidents over the last seven years has increased the academic community's awareness of the importance of chemical risk assessment in laboratories, both teaching and research. These accidents are described in two reports from the US Chemical Safety Board[1,2] addressing chemical safety in education settings and research labs. Approaches to laboratory chemical risk assessment have been suggested by the American Chemical Society's Committee on Chemical Safety.[3] Many different chemicals are used in laboratory work and these risk assessments require ready access to safety information for a large number of chemicals.

In order to support this risk assessment process by providing researchers, educators and students with electronic access to chemical health and safety information, PubChem[4-8] (https://pubchem.ncbi.nlm.nih.gov) provides a "Laboratory Chemical Safety Summary" (LCSS) data view (**Figure 1**) for thousands of chemicals (https://pubchem.ncbi.nlm.nih.gov/lcss/). This format is intended to supplement the more traditional MSDS by providing a concise view of data types most relevant to use of chemicals as described in the OSHA lab standard.[9]

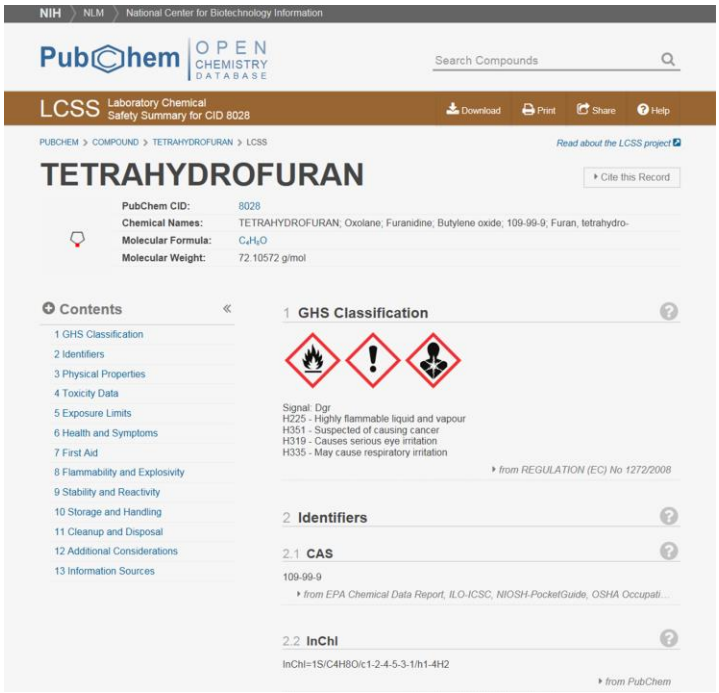

Figure 1. *A partial screenshot of the Laboratory Chemical Safety Summary (LCSS) page for tetrahydrofuran (CID 8028) (https://pubchem.ncbi.nlm.nih.gov/compound/8028#datasheet=lcss).*

PubChem is a chemical information resource developed and maintained by the U.S. National Institutes of Health. PubChem consists of three primary databases: Substance, Compound, and BioAssay. The Substance database (https://www.ncbi.nlm.nih.gov/pcsubstance) stores chemical substance descriptions submitted by individual data contributors. The Compound database (https://www.ncbi.nlm.nih.gov/pccompound) contains unique chemical structures extracted from the Substance database. The BioAssay database (https://www.ncbi.nlm.nih.gov/pcassay) archives descriptions and results of biological assay experiments. Records are cross-mapped between the three databases as described in more detail elsewhere.[4,5]

This conference paper describes PubChem's LCSS data view, including compound coverage, data contents, web accessibility, and download-ability. This paper also describes ongoing community efforts to apply the LCSS to build chemical safety information systems at academic institutions.

PUBCHEM LCSS COVERAGE AND CONTENT

LCSSs are available for PubChem Compound records that have hazard classification information based on the Globally Harmonized System of Classification and Labeling of Chemicals (GHS)[10]. This system is an international standard that establishes technical criteria for classifying chemicals according to their health, physical, and environmental hazards (**Figure 2**). In addition, GHS defines hazard communication requirements for **product labels** and **safety data sheets**. Thus, the GHS provides a global basis for the harmonization of rules and regulations on hazardous chemicals, thereby improving regulatory efficiency and facilitating international trade. The GHS is still in the "roll-out" stage in the U.S. and thus some chemical suppliers have not fully updated their information to meet GHS requirements.

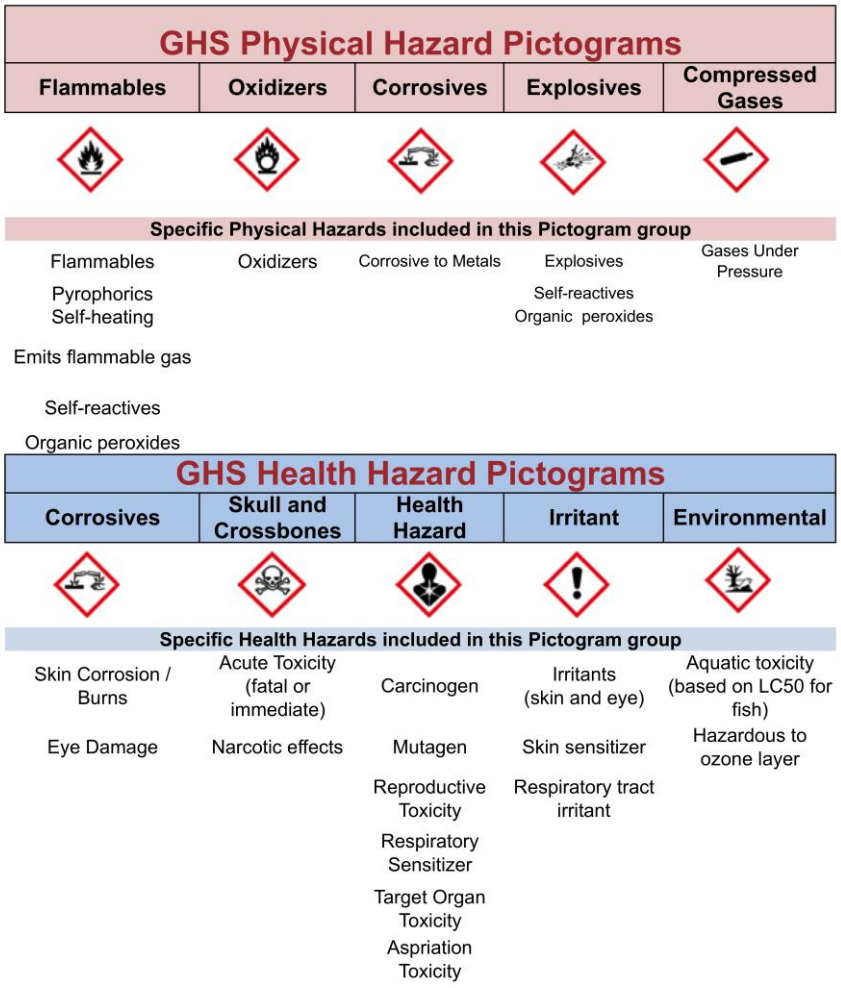

Figure 2. GHS pictograms and hazard classes (provided by Ralph Stuart)

As of October 2015, PubChem has collected GHS hazard classification information for more than three thousand compounds, primarily from two data sources: (1) the International Chemical Safety Cards (ICSC) database[11] at the International Labour Organization (ILO) and (2) Regulation (EC) No. 1272/2008[12,13] (a European Union regulation for classification, labelling and packaging of substances and mixtures). A third source of GHS classification information from the National Institute of Technology and Evaluation in Japan is being processed at the time of this writing.

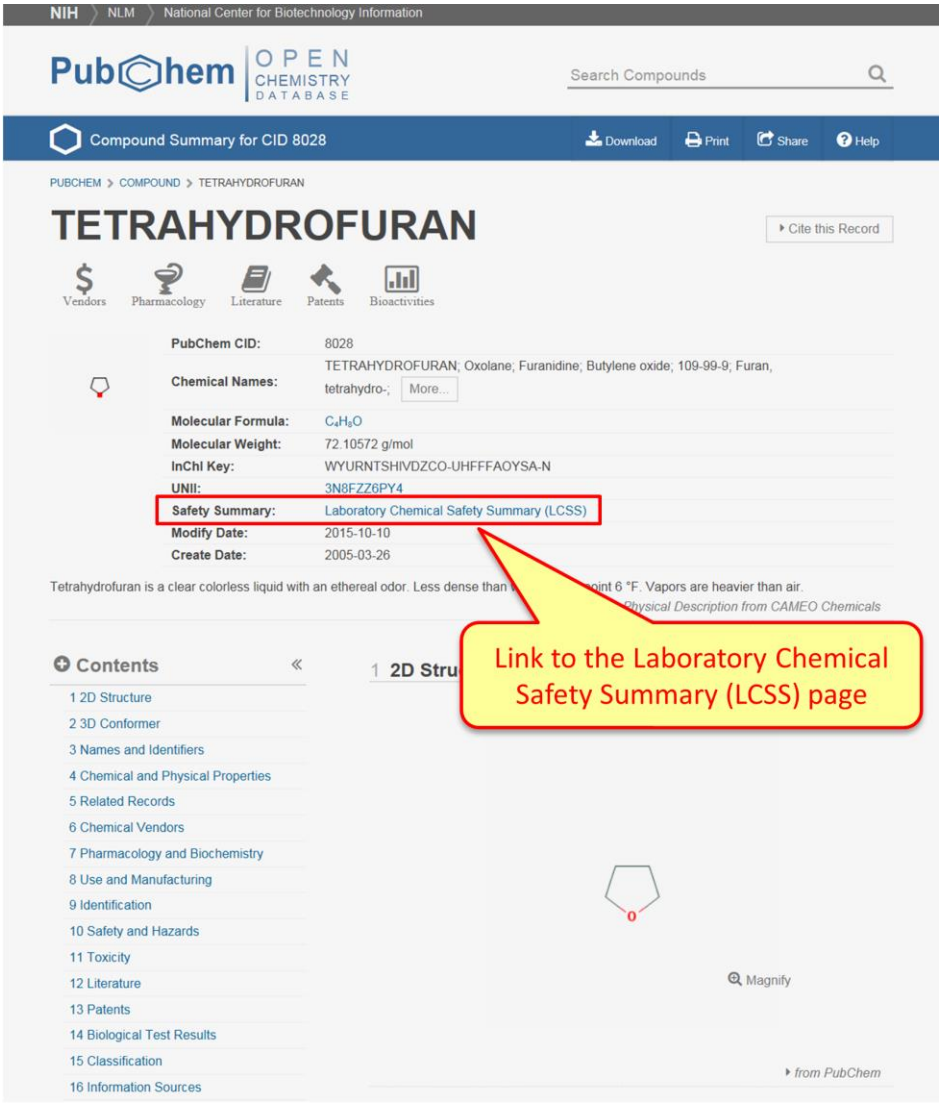

Figure 3. A partial screenshot of the Compound Summary page for tetrahydrofuran (CID 8028) (https://pubchem.ncbi.nlm.nih.gov/compound/8028). If a compound has hazard classification information based on the Globally harmonized system of Classification and Labeling of Chemicals (GHS), the Compound Summary page for the compound has a link to the laboratory Chemical Safety Summary page under the heading "Safety Summary".

When a PubChem Compound record has an LCSS, the link to view it is provided at the top of the Compound Summary page of that compound under the heading "Safety Summary" (**Figure 3**). One can obtain the complete list of chemicals with an LCSS by

visiting the PubChem LCSS project webpage (https://pubchem.ncbi.nlm.nih.gov/lcss/) or by using the PubChem Classification Browser (https://pubchem.ncbi.nlm.nih.gov/classification/) (**Figure 4**).

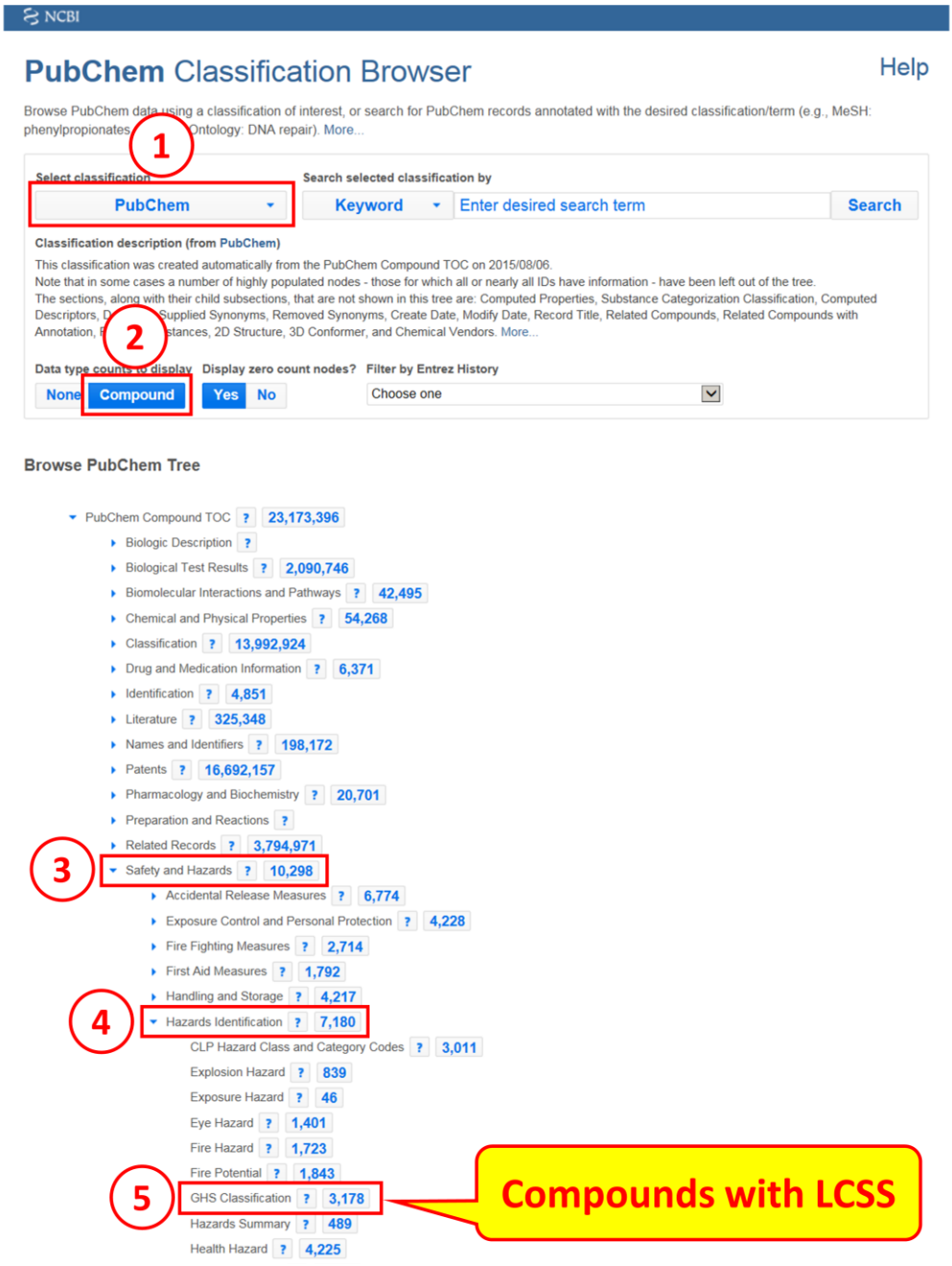

Figure 4. A partial screenshot of the classification browser (https://pubchem.ncbi.nlm.nih.gov/classification/). The complete list of compounds with an LCSS page can be obtained by choosing the following options: PubChem Compound Table of Contents (ToC) → Compound → Safety and Hazards → Hazards Identification → GHS Classifications.

The LCSS contains chemical-specific health and safety data for a given PubChem Compound record, based on the format described by the National Research Council in "*Prudent Practices in the Laboratory: Handling and Management of Chemical Hazards*".[14] **Figure 5** shows information contained in the LCSS and its data sources. The LCSS includes information on flammability, toxicity, exposure limits, exposure symptoms, first aid, handling, clean up and more. These data are provided by various data contributors, including authoritative agencies such as the U.S. National Institute for Occupational Safety and Health (NIOSH), the Occupational Safety and Health Administration (OSHA), the Environmental Protection Agency (EPA), the National Oceanic and Atmospheric Administration (NOAA), and the ILO.

Figure 5. Data contents included in the PubChem Laboratory Chemical Safety Summary (LCSS) at the time of writing.

Of special interest in the PubChem LCSS collection is specific "Stability and Reactivity" information. These entries list hazardous chemical reactions that have been reported between the main chemical in the record and other chemicals and chemical classes. This information is of varied specificity, but often extends beyond that found on Safety Data Sheets (SDS) for the chemical of interest.

VIEWING AND USING DATA IN PUBCHEM LCSS

No one source contains all relevant safety data reported for a chemical and the compiled view of data from several sources provided by the LCSS is convenient and reduces time searching. However, there are many caveats to working with chemical data from any source, especially in an aggregated system where the information is usually cited as in the original source. It is important for the user to verify that the data they are viewing pertains to the form of the chemical that they care about. They may need to review multiple sources to ascertain the extent of data variability.

Chemical information notation varies among sources, which complicates further validation and analysis of the data. For example with names or name-structure associations: general entries on acetic acid may be further specified in some sources as the tri-hydrate form; the structure for formaldehyde may retrieve data about the gas form as well as the saturated water solution, formalin, which may or may not also contain methanol as a stabilizer. Units, conversions and conditions for measurement also vary widely, such as boiling points reported in deg C, or ºF, at mmHg or atm, etc. All of these variables impact the utility of the data for any purpose.

Direct access to the source of each data entry in the PubChem LCSS is provided by a link next to the data presented. This makes establishing the credibility of the data easy. This feature is important because PubChem only aggregates and reports the information from its sources and does not check accuracy of most properties. For example, for some chemicals, different flashpoints are reported from different sources. These differences may represent different protocols for determining the property at hand, or they may represent errors in the data collection, or perhaps just a typo in the original source. It is the responsibility of the user of the data to make an assessment of whether the data presented is of sufficient accuracy and appropriate for their purpose.

LCSS data can also be downloaded helping to enable additional annotation of the data to specific procedures in place at an institution. Pilot projects for using LCSS data to improve local chemical inventory systems are underway at several campuses in the U.S. There are many options for accessing the data, including downloading records from the LCSS page for each compound in several formats (JSON, XML, etc.), or in bulk as an ftp from the PubChem LCSS project webpage. Additional relevant information contained in the PubChem Compound database, such as molecular weight, molecular formula, etc., can be accessed programmatically through a Representational State Transfer (REST)-like interface called PUG-REST[15], allowing one to automate the local LCSS annotation process with additional data from PubChem.

FUTURE DIRECTIONS

The most immediate priority for improving the PubChem LCSS collection is increasing the number of chemicals covered. This will be driven by identifying sources of GHS classification data for chemicals. The most likely sources for such data are chemical suppliers, particularly for those chemical suppliers who specialize in laboratory chemicals. Supplier generated safety sheets may not cover all data types relevant to laboratory use as part of the routine process of bringing a chemical to market. For this reason, it will also be important to continue expanding the number and diversity of additional authoritative public data sources (for example, government agencies and

professional organizations similar to those currently represented in the database) to further enrich the LCSS data view.

A second area of development is looking into opportunities to improve the annotation and classification of the health and safety data in PubChem LCSS. Currently much of the information valuable for risk assessments is locked in strings of text from the original sources. Efforts are underway to develop vocabularies for richer indexing of entries by relevant safety concepts, such as exposure routes and incompatible storage conditions. Chemical structure classifications are in development to support reactivity fingerprinting and predicted incompatibilities.

The third area for improving PubChem LCSS usability is the development of documentation and materials that describe how to most effectively use the information available. Supporting these services is an ongoing collaborative effort among academic librarians, educators, health and safety professionals and information scientists. Such collaborations are actively underway in the work of several professional organizations, particularly the technical divisions of the American Chemical Society that represent those stakeholders. The collaboration of these professional communities can help make the PubChem LCSS data an important asset to support laboratory safety in the 21st Century.

DISCLAIMER

The LCSS provided by PubChem is intended to augment, not replace, safe laboratory practices and procedures for chemical information, such as those found in chemical inventory management systems or laboratory-specific personal protective equipment information. It is the responsibility of PubChem users to determine applicability of or gaps in the LCSS information to support safe use of a chemical. In addition, laboratory risks can arise not only from the specific chemicals used, but also from 1) changes in the concentrations and quantities of those chemicals, 2) new chemicals that are produced, 3) energy sources that occur during a laboratory process, among other variables.

Although we are not aware of any limitations or restrictions on the reuse of PubChem LCSS data, we are not able to give unconditional permission for reuse and advise consultation with intellectual property experts when reusing this data. For more information, please see the disclaimer available at https://www.ncbi.nlm.nih.gov/home/about/policies.shtml.

AUTHOR INFORMATION

Corresponding Author: Sunghwan Kim

*E-mail: kimsungh@ncbi.nlm.nih.gov

ACKNOWLEDGMENTS

This work was supported in part by the Intramural Research Program of the National Library of Medicine, National Institutes of Health, U.S. Department of Health and Human Services.

REFERENCES

1. United States Chemical Safety Board. Key Lessons for Preventing Incidents from Flammable Chemicals in Educational Demonstrations. *Safety Bulletin*, Oct. 2014. http://www.csb.gov/assets/1/19/Lab_Safety_Bulletin_2014-10-30.pdf (accessed Oct. 21, 2015).

2. United States Chemical Safety Board. *Texas Tech University Laboratory Explasion*. Washington, DC, 2011. http://www.csb.gov/assets/1/19/CSB_Study_TTU_.pdf (accessed Oct. 21, 2015).

3. American Chemical Society, Committee on Chemical Safety, Hazard Identification and Evaluation Task Force. *Identifying and Evaluating Hazards in Research Laboratories*. Washington, DC, 2015. http://www.acs.org/content/dam/acsorg/about/governance/committees/chemicalsafety/publications/identifying-and-evaluating-hazards-in-research-laboratories.pdf (accessed Oct. 21, 2015).

4. Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H., PubChem Substance and Compound databases. *Nucleic Acids Res.* **2015**, doi: 10.1093/nar/gkv951.

5. Wang, Y. L.; Suzek, T.; Zhang, J.; Wang, J. Y.; He, S. Q.; Cheng, T. J.; Shoemaker, B. A.; Gindulyte, A.; Bryant, S. H., PubChem BioAssay: 2014 update. *Nucleic Acids Res.* **2014**, 42, D1075-D1082.

6. Wang, Y. L.; Xiao, J. W.; Suzek, T. O.; Zhang, J.; Wang, J. Y.; Bryant, S. H., PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* **2009**, 37, W623-W633.

7. Agarwala, R.; Barrett, T.; Beck, J.; Benson, D. A.; Bollin, C.; Bolton, E.; Bourexis, D.; Brister, J. R.; Bryant, S. H.; Canese, K.; Clark, K.; DiCuccio, M.; Dondoshansky, I.; Federhen, S.; Feolo, M.; Funk, K.; Geer, L. Y.; Gorelenkov, V.; Hoeppner, M.; Holmes, B.; Johnson, M.; Khotomlianski, V. E.; Kimchi, A.; Kimelman, M.; Kitts, P.; Klimke, W.; Krasnov, S.; Kuznetsov, A.; Landrum, M. J.; Landsman, D.; Lee, J. M.; Lipman, D. J.; Lu, Z. Y.; Madden, T. L.; Madej, T.; Marchler-Bauer, A.; Karsch-Mizrachi, I.; Murphy, T.; Orris, R.; Ostell, J.; O'Sullivan, C.; Panchenko, A.; Phan, L.; Preuss, D.; Pruitt, K. D.; Rubinstein, W.; Sayers, E. W.; Schneider, V.; Schuler, G. D.; Sherry, S. T.; Sirotkin, K.; Siyan, K.; Slotta, D.; Soboleva, A.; Soussov, V.; Starchenko, G.; Tatusova, T. A.; Trawick, B. W.; Vakatov, D.; Wang, Y. L.; Ward, M.; Wilbur, W. J.; Yaschenko, E.; Zbicz, K.; Coordinators, N. R., Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2015**, 43, D6-D17.

8. Bolton, E. E.; Wang, Y.; Thiessen, P. A.; Bryant, S. H. PubChem: integrated platform of small molecules and biological activities. In *Annual Reports in Computational Chemistry*, Wheeler, R. A.; Spellmeyer, D. C., Eds.; Elsevier: 2008; Vol. 4, Chapter 12, pp 217-241.

9. *Toxic and Hazardous Substances: National Research Council Recommendations Concerning Chemical Hygiene in Laboratories (Non-Mandatory)*. 29 CFR $1910.1450 Appendix A. Occupational Safety and Health Administration, Washington, DC, 2012. https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=STANDARDS&p_id=10107 (accessed Oct. 21, 2015).

10. About the Globally Harmonized System of Classification and Labelling of Chemicals (GHS). http://www.unece.org/trans/danger/publi/ghs/ghs_welcome_e.html (accessed Oct. 21, 2015).

11. International Chemical Safety Cards (ICSC) database. http://www.ilo.org/safework/info/publications/WCMS_113134/lang--en/index.htm (accessed Oct. 21, 2015).

12. Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No

1907/2006. http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32008R1272 (accessed Oct. 21, 2015).

13. GHS/CLP: Regulation (EC) No 1272/2008 (CLP) - REACH Compliance -. http://www.reach-compliance.eu/english/compliance/CLP/CLP-Initial-Version.html (accessed Oct. 25, 2015).

14. National Research Council, *Prudent Practices in the Laboratory: Handling and Management of Chemical Hazards, Updated Version*. The National Academies Press: Washington, DC, 2011.

15. Kim, S.; Thiessen, P. A.; Bolton, E. E.; Bryant, S. H., PUG-SOAP and PUG-REST: web services for programmatic access to chemical information in PubChem. *Nucleic Acids Res* **2015**, 43, W605-11.